**3GPP TSG-SA WG4 Meeting #130S4-241938**

**USA, Orlando, 18 – 22 November 2024**

**Source: Huawei, HiSilicon**

**Title: [FS\_AMD] WT#12: Summary and conclusions.**

**Spec: TR 26.804**

**Agenda item: 8.6**

**Document for: Agreement**

**1. Introduction**

In order to move forward the WT#12 for FS\_AMD, this paper intends to provide the summary and conclusions to enable the normative work.

**2. Proposal**

It is proposed to agree the following changes to TR 26.804 **CR0007**:

1. Removing editor’s notes, or converting them into NOTEs.

2. Addition of text to summary and conclusions clause.

\* \* \* \* First change \* \* \* \*

##### 5.23.1.2.1 Support of ECN marking for L4S

As described in RFC 9330 [X1], RFC 9331 [X2] and RFC 9332 [X3], the purpose of ECN marking for L4S (Low Latency, Low Loss and Scalable Throughput) is to inform a recipient host at the earliest opportunity that an IP packet has experienced network congestion at some point in its routing path. It exposes congestion information by marking ECN bits in the IP header of the user IP packets between the UE and the application server. This early notification may be used by the receiving application to report the congestion to its sending peer using a suitable Layer 4 feedback mechanism. Based on this feedback, the sender should reduce the sending bit rate. In the context of adaptive segmented media delivery, application layer rate adaptation may be needed in addition. For example, a media player consuming a media presentation that receives an ECN-marked downlink packet from a streaming media server may induce a reduction in the sender’s bit rate by switching to a representation of a lower bit rate. To support this functionality, the recipient host needs to support L4S feedback as described in RFC 9330 [X1].

L4S is based on the idea that delay is mainly caused by the classic congestion control algorithms introduced with TCP. L4S replaces these traditional congestion control algorithms with a class of scalable congestion control algorithms.

The L4S architecture relies on 3 components to operate:

- A scalable congestion control algorithm,

- A modified ECN marking behaviour,

- An active queue management algorithm that isolates L4S traffic

An example of a scalable congestion control algorithm that is widely deployed today is TCP Prague. In TCP Prague the congestion window is adjusted proportionally to the probability of receiving an ECN mark.

$cwnd= \frac{1}{p^{α}}$ , where p is the probability of receiving an ECN mark, and $α$ is a constant

TCP Prague further adjust the congestion window to implement an RTT independence, which is crucial to not overly react to RTT variations. The additive increase part of the algorithm is also adjusted to ensure that flows with short RTT are not penalized by a slow increase.

Another popular scalable congestion control algorithm is BBRv2, which stands for Bottleneck Bandwidth and Round-trip propagation time. BBRv2 continuously estimates the bottleneck bandwidth of the connection and the RTT. It then uses the 2 parameters to adjust its congestion window as follows:

$cwnd=BtlBW\*RTprop$ , where BtlBW is the estimated bottleneck bandwidth and RTprop is the estimated minimum round-trip propagation time. The algorithm’s reaction to packet loss and ECN marks is also modified to be less aggressive, compared to classic congestion control algorithms. To accurately estimate the bottleneck bandwidth, BBRv2 periodically sends data at a high rate (thank the estimated bottleneck bandwidth) to probe the network for the maximum throughput. BBRv2 also tracks the data that is sent and received during a RTT cycle.

The ECN mechanism is also modified in L4S. The default reaction to an ECN mark is to treat it as an equivalent to a packet loss. This is no more the case in L4S, which allows the network to signal the start of congestion more frequently without the risk of messing up the connection’s bandwidth. This change requires the identification of L4S connections by the network. The ECT code point is used for this purpose. When set to 1, the network identifies the sender as L4S capable and deploys the fine-tuned congestion notification mechanism.

Finally, the queue management approach closes the loop for L4S. RFC9332 [X3] defines one such management approach, the Dual Queue Coupled Active Queue Management, which in essence separates the L4S and non-L4S traffic into two separate queues for differentiated latency treatment and ECN marking. The available bandwidth is still shared among both queues though through the coupling mechanism.

According to clause 6.1.3.22 of TS 23.503 [41], an Application Function may provide an explicit indication that the uplink and/or downlink path of a service data flow supports ECN marking for L4S by means of the Nnef\_AfsessionWithQoS service at reference point N33 or the Npcf\_PolicyAuthorization service at reference point N5. The indication is supported at MediaComponent and MediaSubComponent levels, which provides high flexibility on its usage. Based on AF input and/or local configuration, the PCF indicates to the SMF that ECN marking for L4S is enabled for that service data flow. The SMF accordingly configures ECN marking for the corresponding QoS Flow in the uplink and/or downlink direction. ECN marking for L4S in the IP header is supported in either the NG-RAN (see clause 5.37.3.2 and TS 38.300 [X4]), or in the PDU Session Anchor (PSA) UPF (see clause 5.37.3.3 of TS 23.501[23]).

In the case of ECN marking for L4S by the PSA UPF, the NG-RAN is instructed to perform congestion information monitoring and report to the PSA UPF the congestion information of the QoS Flow in the uplink and/or downlink directions via GTP-U header extension. Accordingly, the PSA UPF may mark the uplink and/or downlink direction packets as congested.

\* \* \* \* Next change \* \* \* \*

#### 5.23.2.1 General

Collaboration scenarios 2–11 and 13–15 from TS 26.501 [15] are potential points of departure for improved QoS handling support with the following additions:

1. Similar to the Network Assistance feature in TS 26.501 [15], the network status of the 5G System may be exposed to media delivery sessions using the *QoS monitoring* feature and the *ECN marking for L4S* feature. The network status, including the data rate, latency, congestion, etc. may be used by the Media Delivery System for bit rate adaptation and/or congestion control.

 The PDU Set handling feature may be used to label PDUs belonging to a video frame or video slice as members of the same PDU Set.

NOTE: Whether the concept of PDU Set is feasible for video segment in a segment-based streaming service is not clear.

2. In the case of network congestion, the NG-RAN may consider the PDU Set Importance for PDU Set level packet discarding. This is not expected for segment-based devliery where the TCP or QUIC transport connection used to carry the media streaming service requests reliable transmission.

NOTE: Whether PDU Set feature is beneficial for Media Streaming services is for future study.

\* \* \* \* Next change \* \* \* \*

#### 5.23.4.1 Integrating QoS monitoring and/or ECN marking for L4S

The high-level call flow for integrating the QoS monitoring and/or ECN marking for L4S is shown below as well as the corresponding procedures.

It is assumed that the MNO and the 5GMS Application Provider have negotiated a Service Level Agreement that allows the 5GMS Application Provider to enable the ECN marking for L4S and QoS monitoring in the 5G System for media delivery.



Figure 5.23.4.1-1: Potential call flow for improved QoS handling support

Prerequisites:

- The 5GMS Application Provider has agreed an SLA with the Network Operator to allow the usage of network assistance for Media Streaming service.

Steps:

1. The 5GMS Application Provider provisions the 5G Media Streaming System configures content ingest. **A Network Assistance configuration is provided to allow the usage of ECN marking for L4S and/or QoS monitoring to notify the 5GMS Client of the latest network status.**

2. When the 5GMS-Aware Application starts, the Media Session Handler retrieves the Service Access Informaiton via M5 or M8. The 5GMS AF address that offers the network assistance is provided in the Service Access Information **and the options for QoS monitoring and/or ECN marking are also present**.

3. The Media Session Handler invokes the **Enhanced** Network Assistance API on the 5GMS AF **to find out about the latest network status. For instance, the 5GMS Media Session Handler may subscribe to the periodic congestion status report from 5GMS AF.**

4. The 5GMS AF interacts with the PCF or NEF to enable QoS monitoring and/or ECN marking for L4S in the 5G System via the Npcf\_PolicyAuthorization service at reference point N5 or the Nnef\_AFsessionWithQoS service at reference point N33.

5. In the case of QoS monitoring, the 5GMS AF can receive the notifications from PCF or NEF via the Npcf\_PolicyAuthorization\_Notify at reference point N5 or the Nnef\_AFsessionWithQoS\_Notify at reference point N33.

6. Alternatively, in the case of QoS monitoring, the 5GMS AF may receive the notifications directly from the UPF via the Nupf\_EventExposure\_Notify or from NEF the Nnef\_EventExposure\_Notify service at reference point N33. This is beneficial when the 5GMS AF is deployed in the Edge DN and the SMF/PCF is generally deployed centrally.

7. The 5GMS AF further sends the notifications exposed by the network to the Media Session Handler using the MQTT notification channel for the Provisioning Session. The 5GMS Client may take this into account for rate adaptaion, congestion/flow control.

\* \* \* \* Next change \* \* \* \*

### 5.23.5 Gap analysis and requirements

\* \* \* \* Next change \* \* \* \*

#### 5.23.5.2 QoS monitoring for media streaming

Based on the call flow in clause 5.23.4.3, the following observations are made:

- The 5GMS AF needs to explicitly request QoS monitoring by the 5G System for specific parameters (i.e., congestion information, packet latency, data rate and Packet Delay Variation) by interacting with the PCF at reference point N5 (or else via the NEF at reference poiont N33).

- The Policy Template resource structure at reference point M1 needs to be extended to include the QoS monitoring configuration, including the parameters to be monitored, reporting frequency (event triggered, periodic), optionally the target entity of reporting and optionally the notification via UPF.

- The Policy Template Binding data structure carried in the Service Access Information resource at reference point M5 needs to be extended to reflect the QoS monitoring configuration in the corresponding Policy Template.

- QoS monitoring results need to be exposed to the 5GMS AF, either at reference point N5 using the Npcf\_PolicyAuthorization\_Notify service, or directly with a UPF using the Nupf\_EventExposure\_Notifyservice, or else via a NEF using the Nnef\_EventExposure\_Notifyservice at reference point N33.

- To expose QoS monitoring results to the Media Session Handler in the 5GMS Client, notification events relating to Dynamic Policies at reference point M5 need to be extended to include the QoS monitoring results.

- The QoS monitoring results need to be further provided to the Media Stream Handler by the Media Session Handler at reference point M11.

\* \* \* \* Next change \* \* \* \*

### 5.23.6 Candidate solutions

\* \* \* \* Next change \* \* \* \*

### 5.23.7 Summary and conclusions

This Key Issue has explored several QoS features which could be beneficial to the Media Delivery System, including:

1. ECN marking for L4S,

2. PDU Set handling and

3. QoS monitoring.

Detailed description of the above features is provided in clause 5.23.1.2. Moreover, the high-level call flows for features 1 and 3 are documented in clause 5.23.4, and corresponding gaps are analysed in clause 5.23.5. Finally, candidate solutions to address the identified gaps are provided in clause 5.23.6.

Based on the study, it is recommended to make the following changes to TS 26.501 [15]:

1. Integrate *ECN marking for L4S* into the architectures, high-level call flows and collaboration scenarios for both 5GMSd and 5GMSu.

1. Dynamic policies need to be extended to include the L4S enablement flag in clause 4.0.6, including the domain model and corresponding description.

2. New call flows on the ECN marking for L4S in clause 5.7 and clause 6.9 for downlink and uplink Media Streaming separately.

3. Informative collaboration scenarios for ECN marking for L4S need to be added in Annex A.

2. Integrate the *QoS monitoring* feature into the architectures, high-level call flows and collaboration scenarios for both 5GMSd and 5GMSu.

1. Dynamic policies need to be extended to include the QoS monitoring configuration in clause 4.0.6, including the domain model and corresponding description.

2. New call flows on the QoS monitoring in clause 5.7 and clause 6.9 for downlink and uplink Media Streaming separately.

Based on the study, it is recommended to make the following changes to TS 26.510 [26510]:

1. Integrate *ECN marking for L4S* into the Media Delivery System to support downlink and uplink 5G Media Streaming.

a. Enhance the Dynamic Policy provisioning to include an L4S enablement flag in the Policy Template in clause 5.2.7.

b. Extend the Dynamic Policy invocation to indicate the activation of ECN marking for L4S by Media Session Handler in clause 5.3.3.

c. Enhance policy control interactions for Dynamic Polices to support the request to 5GS (i.e. PCF, NEF) for ECN marking for L4S in clause 5.5.3 and 5.4.3.

d. Extend the *PolicyTemplate* resource to include the L4S enablement flag in clause 8.7.3.1.

2. Integrate the *QoS monitoring* feature into the Media Delivery System to support 5G Media Streaming.

a. Enhance the Dynamic Policy provisioning to include the QoS monitoring configuration in the Policy Template in clause 5.2.7.

b. Extend the Dynamic Policy invocation to indicate the activation of QoS monitoring configuration by Media Session Handler in clause 5.3.3 and 5.4.3.

c. Enhance policy control interactions for Dynamic Polices to support the request to 5GS (i.e. PCF, NEF) for QoS monitoring in clause 5.5.3.

d. Extend the *PolicyTemplate* resource to include the QoS monitoring configuration in clause 8.7.3.1.

e. Extend the DynamicPolicy resource to include the QoS monitoring results in clause 9.3.3.1.

f. Extend the Dynamic Policy client API to enable the Media Session Handler to expose the QoS monitoring results at reference points M6/M11 in clause 11.3.

\* \* \* End of Changes \* \* \* \*